DOCUMENT RESUME

ED 045 692 TM 000 145

AUTHOR	Mayeske, George W.; And Others
TITLE	A Study of Our Nation's Students: Background, Objectives, Design and Methodology of the Study.
INSTITUTION	Office of Program Planning and Evaluation (DHEW/OE), Washington, D.C. Division of Elementary and Secondary Programs.
REPORT NO	TP-7
PUB DATE	30 Jan 70
NOTE	37p.
EDRS PRICE	EDRS Price MF-$0.25 HC-$1.95
DESCRIPTORS	Academic Achievement, Data Analysis, *Equal Education, Ethnic Grouping, Family Background, *Models, Research Design, *Research Methodology, *Research Projects, Socioeconomic Background, *Student Characteristics, Student Motivation, Students
IDENTIFIERS	Coleman Report, *Equality of Educational Opportunity Survey

ABSTRACT

This paper, number 7 in a series, describes the background for a new study to be based on the Equality of Educational Opportunity Survey data and to be entitled "A Study of Our Nation's Students." The nature and scope of the survey are defined and relevant background and conclusions from the prior reported studies based on this data, "Equality of Educational Opportunity" (The Coleman Report) and "A Study of Our Nation's Schools," are educed. The new study focuses on individual students as the unit of analysis. In particular, the main effects and interactions of family background, school variables, geographic region, racial and ethnic background, and sex differences on achievement and motivation of the student will be studied. The data analysis model, the correlation matrix, and the commonality model to be used in the new study are described. Finally, a description of the individual student, student body, school, and geographic variables and sets of variables to be studied concludes the paper. See also TM 000 144 and TM 000 182. (IR)

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
OFFICE OF EDUCATION
WASHINGTON, D.C. 20202

T,MP

JAN 3 0 1970

OFFICE OF PROGRAM PLANNING AND EVALUATION

MEMORANDUM

TO : Technical Paper Recipients

FROM: George W. Mayeske George W Mayeske
Office of Program Planning and Evaluation

SUBJ: Technical Paper Number 7: Background, Objectives, Design and Methodology of the Study

This paper presents the background, objectives, design and methodology of "A Study of Our Nation's Students". Although not in final form this paper will in time become Chapter 2.0 of the larger report focusing on the student and different aspects of his background. This paper is circulated now in order to solicit reactions to its mode of presentation. A summary is given on the last page.

ED045692

OE/PPE
Working Paper
Does Not Reflect
Official Policy of
The Office of Education

U.S. DEPARTMENT OF HEALTH, EDUCATION
& WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

OFFICE OF PROGRAM PLANNING AND EVALUATION

Division of Elementary and Secondary Programs

Background, Objectives, Design and Methodology of the Study

by

George W. Mayeske
Tetsuo Okada
Albert E. Beaton, Jr.
Wallace M. Cohen
Carl E. Wisler

Technical Paper

Number 7

January 30, 1970

TM 000 145

OFFICE OF EDUCATION/U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

2. Background, Objectives, Design and Methodology of the Study

This chapter describes the background for this study including the nature and scope of the data base used, the objectives and main methodological techniques employed and, includes descriptions of key variables and sets of variables used throughout later chapters.

2.1 The Equality of Educational Opportunity Survey

The Civil Rights Act of 1964 requested the Commissioner of Education to "conduct a survey and make a report to the President and the Congress, within two years of the enactment of this title concerning the lack of availability of equal educational opportunities for individuals by reason of race, color, religion or national orgin in public educational institutions at all levels in the United States, its territories and possessions, and the District of Columbia".

In response to this request the Educational Opportunities Survey was carried out by the National Center for Educational Statistics of the U.S. Office of Education, directed by Alexander M. Mood. In addition to its own staff, the Center used the services of outside consultants and contractors. James S. Coleman of Johns Hopkins University had major responsibility for the design, administration and analysis of the Survey. Earnest Q. Campbell of Vanderbilt University shared this responsibility and, in the case of the college surveys, assumed the greater share of it. Frederic D. Weinfeld served as Project Officer for the Survey.

The Survey addressed itself to four major questions:

(1) To what extent are the racial and ethnic groups segregated from one another in the public schools?

(2) Do the schools offer equal educational opportunities in other respects?

(3) How much can students be said to learn, judged by their performance on standardized achievement tests?

(4) What kinds of relationships may be supposed to exist between a student's achievement and the kind of school he attends?

Work was started on the survey in the spring of 1965 with a view to administering the questionnaires and tests that fall. Approximately 70 percent of the schools that were requested to participate in the study actually did so. This entailed testing and surveying some 650,000 students in approximately 4,000 schools throughout the country in grades 1, 3, 6, 9 and 12, together with their teachers, principals and superintendents.

On the basis of competitive bids, the Educational Testing Service of Princeton, New Jersey, was awarded the contract for conducting the Educational Opportunity Survey, including test administration, test scoring, data processing and data analysis. They also consulted on various aspects of the Survey and convened an Advisory Panel to aid in its design and analysis.

The Survey used a 5% sample of schools. This was a two-stage, self-weighting, stratified cluster sample. The Primary Sampling Units (PSU's) in the first stage were counties and Standard Metropolitan Statistical Areas (SMSA's). The PSU's in the second stage were high schools. When one was drawn in the sample the elementary schools feeding into that school were automatically included in the sample as well. Since the Educational Opportunities Survey was primarily concerned with the children of minority groups, and since these groups constituted only about 10% of the total school population, the schools were stratified according

to their percentage of non-white students. Strata with higher percentages of these students were given larger sampling ratios and thus were sampled more heavily. The final result was that over 40% of the students in the Survey were from minority groups.

Separate questionnaires were administered to teachers, principals, superintendents, and students at each of the grade levels studied. The Teacher Questionnaire contained some 72 questions including:

> personal data, professional training, type of college attended, teaching experience, type of school and student preferred, job satisfaction, opinions on issues and problems of integration (busing, compensatory education, etc.), and problems existing in their school.

The final part of the Teacher Questionnaire consisted of a voluntary test of 30 contextual vocabulary items. The purpose of this test was to get a measure of the teachers' verbal facility.

The 100-item Principal Questionnaire was the main source of information about the school. The questions covered school facilities, staff, programs, racial composition, problems, curricula, extra-curricular activities, and many other school characteristics. There were also questions on the personal background and training of the principal and his opinions on the problems of integration.

The Superintendent Questionnaire consisted of 41 questions. In addition to miscellaneous administrative information about the school system, including its expenditures, the questions dealt with the Superintendent himself and his attitudes toward current educational issues.

Detailed factual and attitudinal data about the students were also obtained by questionnaire. Included were items of home background information so that these data could later be equated for such items as socio-economic status, family background, family interest in education, etc.

Different questionnaires were used for each of the grade levels. The Twelfth-Grade Student Questionnaire for example, was comprised of some 116 items. In addition to the questions on home background and the usual personal and school data there were questions on the student's attitude towards school, race relations, and life in general. Representative examples of each category are: "How good a student do you want to be in school?" "If you could be in the school you wanted, how many of the students would you want to be white?" "Good luck is more important than hard work for success". (Agree or Disagree)."

Tests of the various school skills were to be the yardsticks for measuring the detrimental effects of poor school facilities and characteristics upon student learning. The Test Battery was designed as an integral part of the entire research design. The object was to obtain as much data as possible within the limitations of time and available resources. Two of the basic skills chosen were reading comprehension and mathematics ability. These two areas are common to all school curricula and all grade levels. Another area deemed important was that of the general level of knowledge gained by the students, either from their school courses or from experiences in the outside world. A test of general information was therefore included in the Test Battery. Two other ability tests were used to measure the students' verbal and ratiocinative skills.

One major limitation on the design of the test battery was the time required for test administration. It was considered both desirable and administratively feasible to have the test battery and the questionnaires completed in no more than one school day. The lower grades had to have a shorter battery because of the limited attention span of the younger children. Therefore, the testing time increased in the various test batteries until it reached its maximum length in the 12th grade.

The lead time before the administration of the Survey in September 1965 was too short to develop specific tests in the above areas. For this reason, existing standardized tests were used. However, because full-length standardized tests usually required more time than would have been available, it was decided to use shortened, or half-length, forms of these tests rather than to omit tests in any area. Another administrative requirement was that the various tests be interlocked through as many grades as possible so that scores on the same type of tests given at different grade levels could be compared. The scaling allowed us to have a comparable measure of growth between the different grades.

The Act required that the Survey be made at "all levels." It was therefore decided to administer the tests to selected grades at spaced intervals. The expectation was that this would give a good picture of what was going on in the schools while avoiding the need to test at every grade level. The grades chosen were 1, 3, 6, 9 and 12.

Following this survey a report entitled the Equality of Educational Opportunity (EEO) was submitted to the President and the Congress on July 2, 1966, under the principal authorship of James S. Coleman. This report has

become known as "The Coleman Report"; the reader is referred to it for further details of the study (Coleman, et.al., 1966).

The findings from the EEO report that are of particular relevance to this study are summarized in a very general way below. Since much of the current study is an extension of the EEO report, points at which they can be compared are considered in the body of this report as well as in the final chapter. The EEO report found that:

(1) family background was of great importance for achievement;

(2) the relationship of family background to achievement did not diminish over the years of school;

(3) the effect of variations in school facilities, curriculum and staff upon achievement, that was independent of family background was small;

(4) of the school factors, those that had the greatest influence on achievement, independent of family background, were the teacher's characteristics rather than the facilities and curriculum;

(5) the social composition of the student body was more highly related to achievement, independently of the student's own social background, than was any school factor;

(6) attitudes such as sense of control of the environment,or a belief in the responsiveness of the environment were found to be highly related to achievement but appeared to be little influenced by variations in school characteristics.

In summary, the authors of the EEO report concluded: "That the schools bring little influence to bear on a child's achievement that is independent of his background and general social context; and that this very lack of an independent effect means that the inequalities imposed on children by their home, neighborhood and peer environment are carried along to become the inequalities with which they confront adult life at the end of school. For equality of educational opportunity through the schools must imply a strong effect of schools that is independent of the child's immediate social environment and that strong independent effect is not present in American schools".

2.2 A Study of Our Nation's Schools

The EEO report was the first analysis of this data and was conducted in support of the report mandated by Congress under the Civil Rights Act. It was planned to conduct further analyses of this data, and to accomplish this objective a special analysis group was formed in the National Center for Educational Statistics. The first efforts of this group culminated in a report entitled "A Study of Our Nation's Schools" (hereafter called the School study, see Mayeske, et.al., 1969). The steps involved in data analysis for this report as well as the conclusions are summarized on the pages that follow. Some attention is given to the details of these steps since much of this work is drawn upon heavily in the present study.

The School study addressed itself to the following question: How do the school's characteristics influence such things as the achievement level

of all* the students in the school? However, before an answer could be obtained to this question the following technical problems has to be dealt with:

(i) How could discrete categorical variables such as Father's Occupation, best be scaled so that they could be meaningfully interpreted and related to other variables of interest?

(ii) How could provision be made for nonlinear or curvilinear relationships that might otherwise be obscured in the data?

(iii) How could estimates be made of missing data, particularly when those students who failed to provide an answer to a question were of great interest to the analysis?

(iv) How could the more than 400 variables be reduced so that the task of data processing and analysis could be made less complex?

To perform the kind of analysis we wanted and at the same time resolve the above problems the following logical steps were evolved and translated into the necessary computer programs:

Item analyses: Each questionnaire item was analyzed against one or more variables of interest. In this way not only the percent of respondents choosing each item alternative but also their average score

---

*One of the questions addressed in the current study is: How do the school's characteristics relate to the achievement levels of the different kinds of students they get (e.g. different by virtue of sex, race and ethnicity)?

on the variable(s) of interest could be used as a guide in assigning code or scale values for each alternative. The same was true for the non-respondents. For the students' questionnaire item responses were analyzed against an achievement composite.* For the teachers, questionnaire item responses were analyzed against the number of items that were correct on the teacher's vocabulary test.** Questionnaire item responses for the principal's questionnaire were analyzed against the principal's response to questions concerned with his annual salary, number of students enrolled in the school, the rural-suburban-urban location of the school and the proportion of children in the school from working class families.†

Coding and intercorrelation of variables: An approximate ten percent sample of students was drawn from the student master tapes at each grade level. The variables were then coded and intercorrelated.‡ For the teachers and principals a breakdown into elementary and secondary was made and correlations were computed for each breakdown. The full number of teachers and principals included in the survey were used in these analyses.

---

* See Mayeske, et.al., Technical Note No. 64, in the List of References, p. below.

**See Mayeske, et.al., Technical Note No. 32, in the List of References.

† See Mayeske, et.al., Technical Note No. 62, in the List of References.

‡ The codes used for these analyses as well as the means, standard deviations and intercorrelations for the students, teachers and principals are given in the Appendices of the School study (Mayeske, et.al., 1969). The student items were coded using a technique called criterion scaling which maximizes the linear relationship of the questionnaire items with achievement (see Chapter 20 of this report and Appendix XII of the School study).

Reduction of variables into indices. The intercorrelation matrices for the above steps were subjected to a series of factor analyses* in order to obtain meaningful groupings of the variables, called indices.

Computation of index scores. The weights obtained from the factor analyses were used to compute index scores first by standardizing each variable to a mean of zero and a standard deviation of one, and then by multiplying each variable by its respective weight and finally summing these values. In this step index scores were computed for all of the students included in the Survey. Index means, standard deviations and intercorrelations were also computed.**

Computation of school averages. The mean score for each school was computed for both students and teachers on the indices and variables that were carried along separately.

Merging of school data. The school means for students and teachers were merged with the school data for principals on a single tape (one tape for each of the five grade levels).

Computation of correlations and regressions. A large number of statistical analyses were performed inter-relating the variables. The primary statistical tools used were regression analysis and partition of multiple correlation. As a result, we were able to distinguish between:

---

*Principal Components analyses were used with Varimax rotations of Components having a root of one or greater.

**The items and weights used in forming these indices are given in Chapter 20 of this report.

(i) percent of school outcome associated with the distinguishable influence of the school's characteristics;

(ii) percent of school outcome associated with the distinguishable influence of the student's social background;

(iii) percent of school outcome that could just as well be associated with either one.

The conclusions that resulted from these analyses are stated below as a series of hypotheses. Some of the concepts used in these hypotheses and the methods used in support of these hypotheses are given in later sections of this chapter.*

(1) Very little influence of the schools can be separated from the influence of their students' social backgrounds. Conversely, very little of the influence of the students' social background can be separated from the influence of the schools. As the schools are currently constituted children who:

(i) come from the higher socio-economic strata rather than from the lower socio-economic strata;

(ii) have both parents in the home rather than only one or neither parent in the home;

(iii) are white or Oriental-American rather than Mexican-American, Indian-American, Puerto-Rican or Negro; benefit most from their schooling.

---

* In these hypotheses, for example, the students' social background refers to the set of student body social background variables (SBSB) and the characteristics of the school refers to the comprehensive set of school variables (SCHL) as they have been defined and described in the latter part of this chapter.

(2) Until the twelfth grade, the part of the influence of the student's social background that can be separated out is usually larger than the part of the school's influence that can be separated out. At the twelfth grade, however, the distinguishable influence of the school is greater than the distinguishable influence of the student's social background for most of the motivational and attitutinal outcomes. The opposite is true for achievement at the twelfth grade.

(3) The common influence of the school's characteristics and the student's social background on the attitudinal and motivational outcomes differ for the different grade levels. For Achievement, however, the common influence is consistently larger than either one alone. This common influence increases the longer the students stay in school.

(4) Schools that perform well on one outcome tend also to perform well on other outcomes. These performances tend to facilitate and reinforce one another. For the attitudinal and motivational outcomes a school's generalized favorable performance has a large distinguishable influence. It also has a common influence with the student's social background. For the achievement variable the influence of a generalized favorable performance is manifested in common with the school's characteristics and the student's social background.

(5) The school variables that are most heavily involved in school outcomes are those concerned with actual characteristics of the school's personnel, as distinguished from the school's physical facilities, and pupil programs and policies or even from school personnel expenditures such as teaching salaries. (6) Chief among teachers' characteristics related to school outcomes were those reflecting experience in racially imbalanced educational settings. Most non-white teachers had attended predominantly non-white educational institutions and were teaching predominantly non-white students. Non-white educational settings, it was suggested, tend to have associated with them lower levels of achievement and motivation, as well as less favorable socio-economic and family conditions. The result is less adequate preparation than that received in predominantly white institutions.

As a result of these findings the following recommendations were made: (1) in order to obtain specific answers to questions concerned with the improvement of school outcomes a variety of innovations need to be tried in which the outcomes and the school characteristics are specifiable and manipulable so that the results of the innovation(s) can be clearly ascertained; (2) schools should be managed in terms of explicit objectives and performance criteria should exist so that the extent of accomplishment of these objectives through innovative or regular programs can be ascertained; (3) if the dependence of the schools on the social background of their students were lessened (viz., if student bodies were more balanced in terms of their social background) then educational variables might make a greater relative contribution to the students' achievement and motivation.

2.3 Analyzing Differences Among Students

Unlike the School study (where the school was the unit of analysis) the present study focuses on individual students and the ways in which they may differ from one another. Throughout the ensuing chapters the unit of analysis is almost always the individual student. When possible school influences are studied it is usually how they impinge upon the individual student. The major questions for which answers are sought in this study are:

(1) What roles do different aspects of the students' family background play in the development of his achievement and motivation?

(2) What roles do different aspects of the schools play in the development of individual achievement and motivation when juxtaposed with family background factors?

These questions are explored for students in different geographic regions of the country, for students of different racial and ethnic group membership and for boy-girl differences.

Ideally, we would like to study these questions for the same students as they progress through their years of schooling. As pointed out in section 2.1, however, the data are cross-sectional in nature, viz., they were collected from students at different grade levels at one point in time. Consequently, when time trends are inferred they are usually made with great caution. Although the study is primarily descriptive in nature a number of theoretical propositions are formulated from the results of the analyses.

The ensuing chapters tend to be organized along a general-to-specific continuum. The first few chapters give an overview of family background and school influences from a national viewpoint. The succeeding chapters focus on geographic differences followed by racial-ethnic and sex differences. The final chapters focus on racial-ethnic and sex differences within different geographic locations.

The remainder of this chapter presents a discussion of the methodological techniques employed and of the major variables used in the analyses.

2.4 The Data Analysis Model and Its Properties

The data for this study were obtained by appending to each student the attributes of his school (appropriate for his grade level) as they were developed in the School study. This procedure generates a data matrix which is comparable to the following hypothethical matrix:

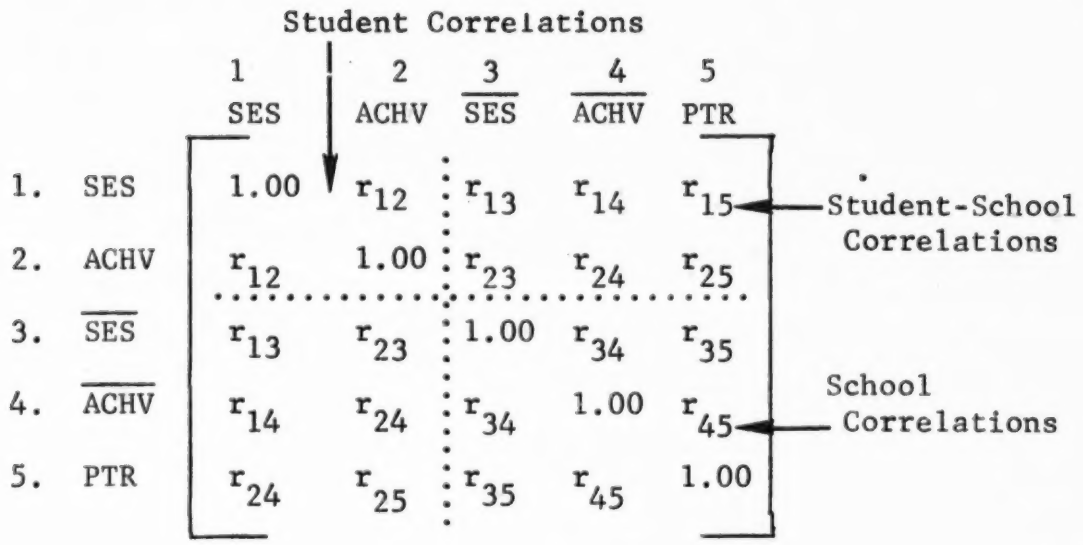
		1 SES	2 ACHV	3 SES	4 $\overline{\text{ACHV}}$	5 PTR
	1	SES_{1A}	$ACHV_{1A}$	$\overline{SES}_A$	$\overline{ACHV}_A$	PTR_A
	2	SES_{2B}	$ACHV_{2B}$	SES_B	$ACHV_B$	PTR_B
Students	3	SES_{3C}	$ACHV_{3C}$	SES_C	$ACHV_C$	PTR_C
	4	.	.	.	.	.
	.	.	.	.	.	.
	.	.	.	.	.	.
	.	.	.	.	.	.
	.	.	.	.	.	.
	N	.	.	.	.	.

In this matrix the individual student, as represented by the numbered rows from 1 to N, is the basic unit of observation. The five columns of the matrix represent different kinds of variables. The first two columns contain observations on each student's Socio-Economic Status (SES) and Achievement (ACHV), respectively. The third and fourth columns contain the average Socio-Economic Status ($\overline{SES}$) and average Achievement ($\overline{ACHV}$) of students in the same school and grade level as the individual student. The last column contains a more traditional school variable, the Pupil-Teacher Ratio (PTR) for students of the same school and grade level. The alphabetical subscripts (e.g. A, B, C) are used to designate the school.

When these variables are intercorrelated the values for each individual student enter into the computational formula* and will result in the following hypothetical correlation matrix:

---

* See Chapter 20 for a comparison of the results obtained when relationships among school variables are computed using the student versus the school as the unit of observation and analysis.



Since this matrix is symmetric, the values below the main diagonal running from the upper left to the lower right will be the same as those above the diagonal. The dotted line is used to separate the submatrix of student correlations from school correlations. Using this matrix and assuming that we are interested in the regression of ACHV on SES we can conduct the following analyses:

TOTAL: the regression of individual student ACHV on SES is given by r^2_{12}. For more than one variable it would be the squared multiple correlation obtained by regressing individual ACHV on several other individual student variables. School variables can be brought into this analysis as well. For example, PTR can be brought into the analysis with SES and ACHV to give " multiple regression of ACHV on PTR and SES.

AMONG: the regression of school $\overline{ACHV}$ on school $\overline{SES}$ is given by r^2_{34}. For more than one variable it would be the squared multiple correlation obtained by regressing school ACHV on several other school variables. Individual

student variables are not brought into this kind of an analysis for reasons given below.

WITHIN: a within school regression is conducted by partialing $\overline{\text{ACHV}}$ out of ACHV, using partial correlation techniques, and then regressing ACHV on SES (i.e., observing the squared partial correlation that remains). This partialing operation renders ACHV uncorrelated with, or independent of ACHV and consequently uncorrelated with any other school variables that are correlated with ACHV*.

$\overline{\text{ACHV}}$ is the one school variable that is most similar to or highly correlated with ACHV. The squared correlation of $\overline{\text{ACHV}}$ with ACHV represents the maximum amount of variance in ACHV that can be explained by analyzing differences among schools. Consequently, when it is partialed out of ACHV all of the remaining school variables are uncorrelated with ACHV. The more general statement is that: when an individual student variable is correlated with its school mean counterpart, that correlation is the maximum value that can be obtained by correlating any other variable, or combination of variables, based upon differences among schools with that individual student variable. When the school mean counterpart is partialed out of an individual student variable all of the differences in that variable associated with

---

* An algebraic proof of this assertion is given in Chapter 20.

differences among schools are removed. This is one of the reasons also that an individual student variable is not entered into an AMONG analysis. For the maximum differences among schools in that variable is just as well represented by its school mean counterpart (e.g. at the school level ACHV is just as well represented by $\overline{\text{ACHV}}$).

2.5 The Commonality Model and Its Properties

In the following chapters extensive use is made of a technique called "Commonality Analysis". Basically, what this technique does is to partition the variance in a dependent variable that is predictable from two or more sets of regressor variables (i.e.,those variables against which the dependent variable is regressed) into the proportion that can be uniquely associated with each set and the proportion that is in common to two or more of the sets. The following discussion will focus on the development of the model for two and three sets of variables and then go on to a discussion of the meaning of these results. A mathematical development of the model is given in Chapter 20.

Let us assume that we have two sets of variables, B and S. In the context of the ensuing chapters, B might represent different measures of the students family background, S might represent different measures of the school the student attends, while A might represent the student's achievement. Suppose now that we run a regression and obtain a squared multiple correlation for A against each set of variables, alone and in combination. For two sets of variables we will have three squared multiple correlations: $R^2(B)$; $R^2(S)$ and; $R^2(B,S)$ where the letter(s) in parentheses represent the set(s) that have been entered into the regression. Then the proportion of the squared

multiple correlation that can be uniquely associated with the B and S sets respectively, designated U(B) and U(S), is given by:

(1) $U(B) = R^2(B,S) - R^2(S)$

(2) $U(S) = R^2(B,S) - R^2(B)$

These unique values are called first order commonality coefficients. The proportion of predictable variance that is common to the two sets of variables, called the second order commonality coefficient, is given by:

(3) $C(B,S) = R^2(B,S) - U(B) - U(S)$

This partitioning results in the following additive properties:

$R^2(B) = C(B,S) + U(B)$

$R^2(S) = C(B,S) + U(S)$

that is, the squared multiple correlations for B and S can be expressed as a function of their different orders of commonality coefficients (viz. the common portion plus the unique portion). In the context of the later chapters this kind of an analysis would indicate the extent to which the predictable variance is shared in common by the two sets and the extent to which it can be associated with one or the other of the two sets. The results of these analyses are organized somewhat as follows:

Order of Commonality Coefficients		B 1	S 2
First	U(Xi)	a	b
Second	C(X1X2)	c	c
	R-SQUARE(Xi)	d	e
	R-SQ(X1,X2)	f	f

In this table the first order commonality coefficient or portion uniquely attributable to each set is given in the U(Xi) row where Xi stands for the set contained in each column, represented by B and S, respectively. The second order commonality coefficient is the same for each column as is the R-SQ(X1,X2). The squared multiple correlation for each set, B or S, is given in the row R-SQUARE(Xi). Also, the following empirical values in this table would be additive: a + c = d, b + c = e and a + b + c = f When we perform a unitizing operation on these results the different orders of commonality coefficients sum to 100. This is performed by dividing each of the empirical values in this table by the value for f. Usually only the unitized values for U(Xi) and C(X1X2) are presented.

For the three set case let us designate the third set as O for Other. From entering all the different combinations of sets in the regression we obtain the following squared multiple correlations: $R^2(B)$; $R^2(S)$; $R^2(O)$; $R^2(B,S)$; $R^2(B,O)$: $R^2(SO)$; $R^2(B,S,O)$. Then the first order commonality coefficients are given by:

$$U(B) = R^2(B,S,O) - R^2(S,O)$$

$$U(S) = R^2(B,S,O) - R^2(B,O)$$

$$U(O) = R^2(B,S,O) - R^2(B,S)$$

The second order commonality coefficients are given by:

$$C(B,S) = R^2(B,S,O) - R^2(O) - U(B) - U(S)$$

$$C(B,O) = R^2(B,S,O) - R^2(S) - U(B) - U(O)$$

$$C(SO) = R^2(B,S,O) - R^2(B) - U(S) - U(O)$$

and the third order commonality coefficient, of which there is only one, is given by:

$$C(B,S,O) = R^2(B,S,O) - R^2(B,S) - R^2(B,O) - R^2(S,O) - U(B) - U(S) - U(O)$$

The squared multiple correlation for any single set can then be expressed as a function of its different order commonality coefficients. For example, the squared multiple correlation for the Other set, $R^2(O)$, can be expressed as:

$$R^2(O) = C(B,S,O) + C(B,O) + C(S,O) + U(O)$$

Results of three set commonality analyses are organized somewhat as follows:

Order of Commonality Coefficients		B 1	S 2	O 3
First	U(Xi)	a	b	c
Second	C(X1X2)	d	d	
	C(X1X3)	e		e
	C(X2X3)		f	f
Third	C(X1X2X3)	g	g	g
	R-SQUARE	h	i	j
	R-SQ(X1X2X3)	k	k	k

With three sets there are now three second order commonality coefficients. The additive properties would be:

$a + d + e + g = h$; $b + d + f + g = i$; $c + e + f + g = j$ and;

$a + b + c + d + e + f + g = k$

When these coefficients are divided by R-SQ(X1X2X3), which in the above table has the empirical value of k, they are called "unitized" coefficients. Usually only these "unitized" coefficients are given in the following chapters. Sometimes only the results for a single column are given, e.g., column 3, the O set.

A number of terms are used interchangeably to refer to the first and higher order commonality coefficients in the chapters that follow The first order coefficient is often called the unique role, unique value, unique portion, or independent role. Similarly, the higher order coefficients (those other than the first) are often referred to as the common role, common portion, common value or shared role.

One might ask what meaning can be attributed to these different coefficients. The first order coefficients (the unique portions U(Xi)), represent that portion of the predictable variance that can be uniquely associated with one of the sets. In the strictest sense the higher order coefficients, (those other than the U(Xi)) represent our inability to separate out the functioning of one set from the other.

In making inferences about the possible influence that the different sets of variables have on the dependent variable it would seem that both the unique (first order coefficients) and the common portions (those coefficients higher than the first) could represent influences. The common portions might represent the joint influence of two or more sets of variables or they might represent the fact that the occurrence of one attribute is accompanied by the occurrence of a second attribute. For example, students from the lower socio-economic strata are more likely to have a less intact family structure, to be less well motivated and have lower achievement, etc. This line of reasoning is further reinforced when we recognize that the unique portion for a set of variables, which is usually considered as representing a causal influence, can be moved up to the higher order when a new set of variables is entered into the analysis with it. This occurs for example, when motivational variables are entered into the analysis with Socio-Exonomic Status and Family Structure.

2.6 An Hypothesis Testing Framework for Stratified Regressions

Much of this report is devoted to the systematic study of how the relationships of family background and school factors with achievement and motivation differ for different subgroups of students. For example, differences among these relationships are explored for students in different areas of the country, for students of different racial-ethnic group membership and ,for boy-girl differences. The following framework, described in detail in Chapter 20, was used to systematically test for the extent of these subgroup differences. The technique used is one outlined by Kuh and programmed by Beaton (1964, see also Chapter 20) and is similar to one presented by Wilson and Carry (1969). This sequential procedure, which utilizes various sums of squares and mean squares from a covariance analysis, runs as follows:

H_1: Are the cell (or subgroup) regressions, (including the cell intercepts) similar to the overall regression obtained when all students are combined without regard to their subgroup membership? If this hypothesis is accepted then the sequence is terminated. However, if this hypothesis is rejected then the next hypothesis in the sequence is tested.

H_2: Are the cell slopes or regression weights (excluding the cell intercepts) similar to the overall slope obtained when all students are combined without regard to their subgroup membership? If this hypothesis is rejected then the sequence is terminated. However, if the hypothesis is accepted then two more tests are available for distinguishing between different

kinds of intercepts. Since only the first two hypotheses are of interest in this study the others will not be discussed (although they are presented in Chapter 20). The F statistic is used to determine whether to accept or reject the hypotheses.

If the subgroups are found to be different using this hypothesis testing framework then, comparative commonality analyses are usually run to determine how the relative roles of the sets of variables of interest may change from one group to another.

2.7. Definition and Description of Variables Used

This section gives a detailed description and interpretation of the variables and sets of variables used throughout the study. When indices (weighted composites of other variables) are discussed the weights used to form the indices are given in Chapter 20. The discussion will deal first with individual student variables, then with school level variables and last, with the definition of different sets of variables. An asterisk alongside an index or variable means that it is more adequately measured at the higher grade levels (6, 9 and 12) than at the lower grade levels (1 and 3).

2.7.1 Individual Student Indices and Variables

Socio-Economic Status (SES)*: a student with a high score on this index has parents who come from the higher educational strata, his father is engaged in a professional, managerial, sales or technical job, there are two to three children in the family, about six to ten rooms in their home, they are more likely to reside in the residential area of the city or the suburbs rather

than in the inner city and there are intellectually stimulating materials accessible in the home such as books, magazines, newspapers, television and radio.

Family Structure and Stability (FSS)*: a student with a high score on this index has both parents in the home, his father's earnings are the major source of income, his mother works part-time or not at all and his family has not moved around much.

Racial-Ethnic Group Membership (RETH): a student with a high score on this variable is white, a student with an intermediate score is Oriental-American and a student with a low score is Puerto-Rican, Mexican-American, Indian-American or Negro-American. An individual's score on this variable represents his membership not only in a physical category but in a social category as well. For, in a society that differentiates on the basis of skin color, ones' membership in a particular racial or ethnic group is very much a social category with many behavorial implications.

Expectations for Excellence (EXPTN)*: a student with a high score on this index says that this mother, father and teachers want him to be one of the best students in his class and that he also desires to be one of the best in his class.

Attitude Towards Life (ATTUD)*: a student with a high score on this index feels that: people who accept their condition in life are not necessarily happier; hard work is more important than good

luck for success; when he tries to get ahead he doesn't encounter many obstacles; with a good education he won't have difficulty getting a job; he would not sacrifice anything to get ahead nor does he want to change himself; he does not have difficulty learning nor does he feel that he would do better if his teachers went slower; and, people like him have a chance to be successful.

Educational Plans and Desires (EDPLN)*: a student with a high score on this index says that his parents want him to go to college, he both desires and plans to go to college, aspires to one of the higher occupational levels and feels that he is one of the brighter students in his class.

Study Habits (HBTS)*: a student with a high score on this index has frequent (weekly or more) discussions with his parents about his school work, was read to regularly as a child, spends one to three hours a day studying and one to three hours a day watching TV, would make most any sacrifice to stay in school and has seldom stayed away from school just because he wanted to.

Achievement (ACHV)*: a student with a high score on this index or composite tended to score high on all of the tests that entered into that composite. For all grade levels the tests of Verbal and Non-Verbal Ability were used as part of the composite. In addition, at grades six, nine and twelve, tests of Reading Comprehension and Mathematics Achievement were used and at grades nine and twelve a

test of General Information was included in the composite. In one sense, this inclusion of more tests at the higher grade levels represents the nature of the educational process, where basic skills are acquired in the early years and then other skills and knowledge are acquired through the use of these basic skills. As shown in Chapter 20, these tests at each grade level were sufficiently highly correlated to be included in a single composite.

Kindergarten Attendance (KGTN): a student with a high score attended kindergarten, a student with a low score did not.

Nursery School Attendance (NRSY): a student with a high score attended nursery school before he attended kindergarten, a student with a low score either did not attend nursery school or could not remember if he did or not (this information available only at grades six, three and one).

2.7.2 Student Body Variables

When the values of a variable are averaged for each of the students in a particular grade level of a school, this results in what is called a Student Body variable. The relationship between a Student Body variable and the same variable at the individual level was outlined in section 2.4 in describing properties of the data analysis model. Schools with a high mean or average on a Student Body variable tend to have a larger proportion of students with a high score on that attribute, while schools with a low mean or average tend to have a larger proportion of students with a low score

on that attribute. The meaning of these variables at the individual student level were indicated earlier. The Student Body variables used in this study are:

Student Body Socio-Economic Status

Student Body Family Structure and Stability

Student Body Racial-Ethnic Group Membership

Student Body Expectations for Excellence

Student Body Attitude Towards Life

Student Body Educational Plans and Desires

Student Body Study Habits

Student Body Achievement

2.7.3. School Variables

In this study, to represent attributes of the schools other than Student Body variables, the following comprehensive set of thirty-one indices and variables is used. A description of the meaning of each index and the variables that comprise it is given in Chapter 20. A detailed description of the development of these indices and variables is given in Mayeske, et.al., 1969. The indices and variables are categorized into three subsets of Facilities, Pupil Programs and Policies, and School Personnel and Personnel Expenditures. All but seven of the thirty-one variables are indices. Unlike the earlier individual student variables there were no problems with the adequacy of measurement of the variables at the lower grade levels.

FACILITIES

(1) Plant and Physical Facilities

(2) Instructional Facilities

(3) Pupils Per Room

PUPIL PROGRAMS AND POLICIES

(1) Tracking

(2) Testing

(3) Transfers

(4) Remedial Programs

(5) Free Milk and Lunch Programs

(6) Accreditation

(7) Age of Texts

(8) Availability of Texts

(9) Pupil Teacher Ratio

(10) Enrollment

SCHOOL PERSONNEL AND PERSONNEL EXPENDITURES

(1) Principal's Experience

(2) Principal's Training

(3) Principal's College Attended

(4) Principal's Sex

(5) Principal's Estimate of the School's Reputation

(6) Specialized Staff and Services

(7) Teacher's Experience

(8) Teacher's Training

(9) Teacher's Socio-Economic Background

(10) Teacher's Localism

(11) Teacher's College Attended

(12) Teaching Conditions

(13) Teaching Related Activities

(14) Preference for Student Ability Level

(15) Teacher's Sex

(16) Teacher's Racial-Ethnic Group Membership

(17) Teacher's Vocabulary Score

In the following chapters the above set of variables is referred to as the School set (SCHL).

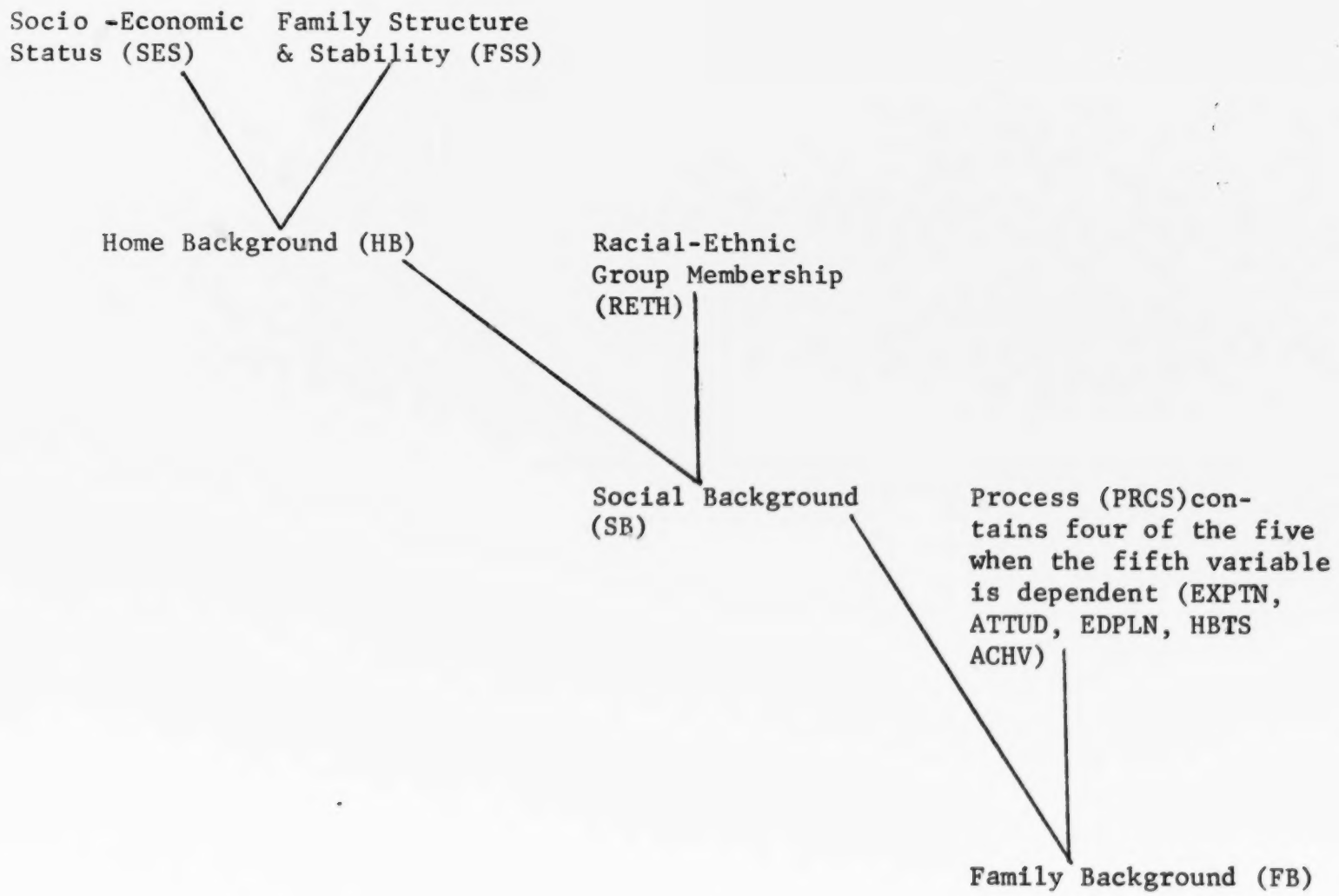
2.7.4 Definition and Description of Sets of Variables

Throughout the ensuing chapters a number of sets of variables are used recurrently. The variables that comprise each of these sets as well as a rationale for their inclusion and a schematic diagram of their interrelationships are given in this section.

HOME BACKGROUND (HB): this label is applied to the set of variables comprised of the student's Socio-Economic Status (SES) and Family Structure and Stability (FSS) since these variables represent the human and material resources in the immediate home environment.

SOCIAL BACKGROUND (SB): this label is applied to the set of HB variables when included with the Racial-Ethnic Group Membership (RETH) variable. This latter variable helps to place the student with regard to the majority-minority aspect of the social structure (or alternatively, the color-caste aspect).

Figure 2.7.4.1. - A Schematic Diagram of the Variables Included in the Different Sets


Socio -Economic Status (SES)
Family Structure & Stability (FSS)
Home Background (HB)
Racial-Ethnic Group Membership (RETH)
Social Background (SB)
Process (PRCS)contains four of the five when the fifth variable is dependent (EXPTN, ATTUD, EDPLN, HBTS ACHV)
Family Background (FB)

FAMILY BACKGROUND (FB): this set is comprised of the SB and PRCS sets; the term family background is used to refer to all aspects of the individual student's background. When separate analyses are run for each racial-ethnic group, FB contains only the HB variables of SES and FSS. The relationships among these sets of variables is given in schematic form in Figure 2.7.4.1.

There are in addition, four sets of variables at the school level. They are:

SCHOOL (SCHL): this is the comprehensive set of thirty-one variables described previously.

STUDENT BODY SOCIAL BACKGROUND (SBSB): this set is comprised of the three Student Body variables of SES, FSS and RETH and, is the counter part of SB at the individual level.

SCHOOL OUTCOMES (SO): this is the set of five Student Body variables of: Expectations; Attitude Towards Life; Educational Plans; Study Habits and; Achievement. This set is called School Outcomes because each of these variables can be regarded as being influenced at least in part, by the school. This set does <u>not</u> vary in its composition depending upon the individual student dependent variables.

FAMILY PROCESS (PRCS): this set is the exact counterpart of PRCS at the individual level and consequently the same name is used. Its composition does vary as follows:

Dependent Variable	Composition of PRCS at School Level The four Student Body Variables of:
Student Body Expectations (EXPTN)	ATTUD, EDPLN, HBTS, ACHV
Student Body Attitude (ATTUD)	EXPTN, EDPLN, HBTS, ACHV
Student Body Ed. Plans (EDPLN)	EXPTN, ATTUD, HBTS, ACHV
Student Body Study Habits (HBTS)	EXPTN, ATTUD, EDPLN, ACHV
Student Body Achievement (ACHV)	EXPTN, ATTUD, EDPLN, HBTS

This latter PRCS set at the school level is usually used only for the AMONG school analyses.

2.7.5 Definition of Geographic Groupings

In a number of chapters comparative analyses are conducted for different geographic groups. The four basic groupings are: Metropolitan (MET); Non-Metropolitan (NON-MET); North and ; South. The standard census tract was used to define metropolitan and non-metropolitan areas as used in the sampling design (discussed insection 2.1). The South was defined to include the 16 Southeastern and Southwestern States of: Alabama, Arkansas, Arizona, Florida, Georgia, Kentucky, Louisiana, Mississippi, New Mexico, North Carolina, Oklahoma, South Carolina, Texas, Tennessee, Virginia and West Virginia. Included as North were the remaining States.

2.8 Summary

This chapter has described the background for this study, including the nature and scope of the Educational Opportunities Survey data which provided the basis for this study. A brief description has been given of two other major studies "Equality of Educational Opportunity" and "A Study of Our Nation's Schools", which have used this data. This chapter also described the data analysis model employed in this study; a technique for the partition of multiple correlation; a framework for systematically testing the differences among regressions for different subgroups; a description of the variables and sets of variables used throughout the study and; a description of the geographic groupings.